

# Conformational transitions in random heteropolymer models

Viktoria Blavatska<sup>1,2,\*</sup> and Wolfhard Janke<sup>1,†</sup>

<sup>1</sup>*Institut für Theoretische Physik and Centre for Theoretical Sciences (NTZ),  
Universität Leipzig, Postfach 100920, D-04009 Leipzig, Germany*

<sup>2</sup>*Institute for Condensed Matter Physics of the National Academy of Sciences of Ukraine,  
79011 Lviv, Ukraine*

We study the conformational transitions in heteropolymers within the frames of a lattice model containing two types of monomers  $A$  and  $B$ . Such a model can describe in particular the sequences of hydrophobic and hydrophilic residues in proteins (K.F. Lau and K.A. Dill, *Macromolecules* **22**, 3986 (1989)) and polyampholytes with oppositely charged groups (Y. Kantor and M. Kardar, *Europhys. Lett.* **28**, 169 (1994)). We generalize this lattice model by introducing various types of short-range monomer-monomer interactions. Applying the pruned-enriched Rosenbluth chain-growth algorithm (PERM) we analyze numerically the peculiarities of transitions from extended into compact states as function of the fraction of  $A$  and  $B$  monomers along the heteropolymer chain.

PACS numbers: 36.20.-r, 82.35.Jk, 07.05.Tp

## I. INTRODUCTION

A subject of great interest in both chemical and biological physics are the conformational properties of long heteropolymer chains. Typical examples are proteins, consisting of sequences of amino acid residues, connected by peptide bonds. The conformations of individual macromolecules are controlled by the type of monomer-monomer interactions. In general, the constituents (monomers) of macromolecules in an aqueous environment can be characterized as hydrophilic or hydrophobic, depending on their chemical structure. Hydrophilic residues tend to form hydrogen bonds with surrounding water molecules, whereas the hydrophobic monomers effectively attract each other and tend to form a dense hydrophobic core. For a flexible polymer chain in an athermal solvent, the steric hard-core repulsions between monomers are dominant, and a typical polymer chain usually attains a coil-like conformation. In this regime, the size measures of a typical  $N$ -monomer chain (such as the mean-square end-to-end distance  $\langle R_e^2 \rangle$  or the mean-square radius of gyration  $\langle R_g^2 \rangle$ ) obey scaling laws [1–3]:

$$\langle R_e^2 \rangle \sim \langle R_g^2 \rangle \sim N^{2\nu_{\text{coil}}},$$

with a universal exponent  $\nu_{\text{coil}} = 0.587597(7)$  in three dimensions [4]. At temperatures below the so-called  $\Theta$ -temperature, growing effect of hydrophobic monomers attraction makes the polymer radius shrink, which causes the collapse transition into the compact globule-like regime. Exactly at the  $\Theta$ -temperature, the macromolecules behave like an ideal Gaussian chain (up to logarithmic corrections) with size exponent  $\nu_{\Theta} = 1/2$ . The conformational properties of such heteropolymers can be successfully studied within the frames of a lattice model, where each monomer is treated as a point particle on a regular lattice. For simplicity, it was proposed to consider lattice polymers which contain only two types of monomers, say  $A$  (hydrophobic) and  $B$  (hydrophilic), the so-called HP model [5]. The conformational state of such a model is defined by the short-range nearest-neighbour attraction between any pair of (non-bonded)  $A$ -monomers along the SAW trajectory, and thus depends on the fraction of residues of the two types and on the sequence of residues along the chain. In further generalizations of this model the contact interaction is replaced by a distance-dependent Lennard-Jones type of potential, being attractive for  $AA$  and  $BB$  pairs and repulsive for  $AB$  pair of monomers [6]. Considerable attention of researchers has been paid to identifying a compact globular state (associated with the global energy minimum) of a chain with fixed  $A$  and  $B$  sequence (describing proteins with a given sequence of amino acid residues) [7–13].

On the other hand, the long-range nature of the electrostatic Coulomb interaction between charged groups also produces crucial effects on heteropolymer conformations. Of particular interest are so-called polyampholytes (PAs) [14, 15]: heteropolymers, comprising both positively and negatively charged monomers. Examples of polyampholytes are synthetic copolymers bearing acidic and basic repeat groups. It is established, that PAs possess a rich conformational

---

\*E-mail: Viktoria.Blavatska@itp.uni-leipzig.de; viktoria@icmp.lviv.ua

†E-mail: Wolfhard.Janke@itp.uni-leipzig.de

behaviour, strongly dependent both on their overall charge  $Q$  and the quality of solvent (and thus on temperature  $T$ ) [16–24]. If positive and negative charges are almost balanced ( $Q$  is small), the attractive Coulomb interaction dominates, and the polymer collapses into a globular (sphere-like) state at low temperatures. If the energy of the electrostatic interactions prevails the globular surface energy, which happens when the excess charge  $Q$  approaches  $Q_c \sim N^{1/2}$ , molecules split into extended necklace-like sequences of small connected globules. In the high-temperature regime but small  $Q$  (predominant electrostatic attraction), the PA forms globule-like structures, whereas at the critical value of the excess charge  $Q_c \sim N^{1/2}$  the charged polymer attains the shape of a neutral polymer coil. Polyampholytes with considerable disbalance between positive and negative charges ( $Q > Q_c$ ) and thus predominant Coulomb repulsion are expected to attain a stretched rod-like configuration. A particular model of PA with strongly screened short-range Coulomb interactions (as proposed in Ref. [25]) appears to be in a close relationship with the HP model as described above; really, in this case one has mutual repulsion between  $AA$  and  $BB$  pairs and attraction for  $AB$  pairs.

In the present work, we aim to study several generalizations of a lattice heteropolymer model with various types of monomer-monomer interactions. In particular, this allows us to compare the properties of HP model and model of a polyampholyte chain with strongly-screened Coulomb interaction directly within the same approach. The main accent will be made on the peculiarities of conformational transitions from extended into compact state as function of the fraction of  $A$  and  $B$  monomers along the heteropolymer chain.

The layout of the rest of the paper is as follows: in the next Section II we introduce the model and briefly describe the pruned-enriched Rosenbluth method, applied in our study. The results obtained are given in Section III. We end up by giving conclusions and an outlook in Section IV.

## II. MODEL AND METHOD

We start with a lattice model of a linear polymer chain with  $N_A$  monomers of type  $A$  and  $N_B = N - N_A$  monomers of type  $B$  in  $d = 3$  dimensions. Denoting by  $\varepsilon$  an interaction energy of any pair of monomers residing on neighboring lattice sites that are not connected by a covalent bond, we consider 5 different cases:

$$\text{Model 1 : } \varepsilon_{AA} = \varepsilon_{BB} = 1, \varepsilon_{AB} = -1, \quad (1)$$

$$\text{Model 2 : } \varepsilon_{AA} = \varepsilon_{BB} = -1, \varepsilon_{AB} = 1, \quad (2)$$

$$\text{Model 3 : } \varepsilon_{AA} = 1, \varepsilon_{BB} = \varepsilon_{AB} = 0, \quad (3)$$

$$\text{Model 4 : } \varepsilon_{AA} = -1, \varepsilon_{BB} = \varepsilon_{AB} = 0, \quad (4)$$

$$\text{Model 5 : } \varepsilon_{AA} = -1, \varepsilon_{BB} = 1, \varepsilon_{AB} = 0. \quad (5)$$

The model 1, where the like monomers repel and the opposite ones attract, refers to strongly screened Coulomb interactions, as studied previously in Ref. [25]. The model 3 is a particular case of model 1 and corresponds to a polymer chain containing charged ( $A$ ) and neutral ( $B$ ) monomers. Model 4 refers to the (minimal) HP model [5] with hydrophobic ( $A$ ) and hydrophilic ( $B$ ) monomers. Cases 2 and 5 can be considered as generalizations of the two above mentioned models.

Let us introduce the inhomogeneity ratio:

$$c \equiv \frac{N_A}{N}, \quad (6)$$

so that the cases  $c = 1$  ( $c = 0$ ) correspond to homogeneous chains with all monomers of type  $A$  ( $B$ ). Note also, that models 1 and 2 are “symmetrical” in an energetical sense with respect to the value  $c = 0.5$ .

To study the conformational properties of models 1-5, we apply the pruned-enriched Rosenbluth method (PERM) [26]. The chain grows step by step, i.e., the  $n$ th monomer is placed at a randomly chosen neighbour site of the last placed  $(n-1)$ th monomer ( $n \leq N$ ). The growth is stopped, if the total length of the chain,  $N$ , is reached (we consider chains of length up to  $N = 100$ ). The total energy of a chain  $E(N)$  is given by

$$E(N) = \sum_{i=1}^N \sum_{j=i+2}^N \varepsilon_{ij} \sigma_{ij}, \quad (7)$$

where each  $i$  and  $j$  is either of the type  $A$  or  $B$ , correspondingly  $\varepsilon_{ij}$  is 1,  $-1$  or 0 depending on the model, and  $\sigma_{ij} = 1$  if monomers  $i$  and  $j$  are nearest neighbours and zero otherwise. The Rosenbluth weight factor  $W_n$  is thus taken to be

$$W_n = \prod_{l=2}^n m_l e^{-\frac{(E(l) - E(l-1))}{k_B T}}, \quad (8)$$

where  $m_l$  is the number of free lattice sites to place the  $l$ th monomer,  $E(l)$  denotes the energy of the  $l$ -step chain, and  $k_B$  is the Boltzmann constant. In what follows, we will assume units in which  $k_B = 1$ .

Pruning and enrichment are performed by choosing thresholds  $W_n^<$  and  $W_n^>$  depending on the current estimate of the sum of weights  $Z_n = \sum_{\text{conf}} W_n^{\text{conf}}$  of the  $n$ -monomer chain [11, 26, 27]. If the current weight  $W_n$  of an  $n$ -monomer chain is less than  $W_n^<$ , the chain is discarded with probability 1/2, whereas if  $W_n$  exceeds  $W_n^>$ , the configuration is doubled (enrichment of the sample with high-weight configurations).

When studying random heteropolymers, a double averaging has to be performed to receive the quantitative value for any observable. At first, the conformational averaging is performed over an ensemble of possible conformations of a macromolecule with a fixed sequence of  $A$  and  $B$  monomers under a given inhomogeneity ratio  $c$  value along the chain:

$$\langle O \rangle = \frac{\sum_{\text{conf}} W_N^{\text{conf}} O}{\sum_{\text{conf}} W_N^{\text{conf}}}, \quad (9)$$

where  $W_N^{\text{conf}}$  is the weight of an  $N$ -monomer chain in a given conformation. The second is the “sequence average”  $\overline{\langle \dots \rangle}$ , which is carried out over different random sequences of  $A$  and  $B$  monomers under given  $c$  value:

$$\overline{\langle O \rangle} = \frac{1}{M} \sum_{i=1}^M \langle O \rangle_i. \quad (10)$$

Here,  $M$  is the number of different sequences and the index  $i$  means that a given quantity is calculated for sequence  $i$ . We consider up to  $M = 1000$  random sequences at given inhomogeneity ratio value. Note, that the case of so-called “quenched disorder” is considered, where the average over different sequences is taken after the configurational average has been performed.

### III. RESULTS

The averaged energy of heteropolymer models 1-5 can be predicted based on the following considerations. Let us introduce the averaged number of nearest-neighbour contacts  $p$  for any monomer in a chain, which can be estimated from the empirical relation:

$$p = 2d - 1 - z, \quad (11)$$

where  $z$  is the fugacity (connectivity constant) of the polymer chain. Having a heterogeneous polymer chain with inhomogeneity ratio  $c$  (so that  $N_A = N \cdot c$ ,  $N_B = N \cdot (1 - c)$ ), we can estimate the number of  $AA$  nearest-neighbour contacts:

$$n_{AA} = \frac{1}{2} N_A p c = \frac{1}{2} N p c^2, \quad (12)$$

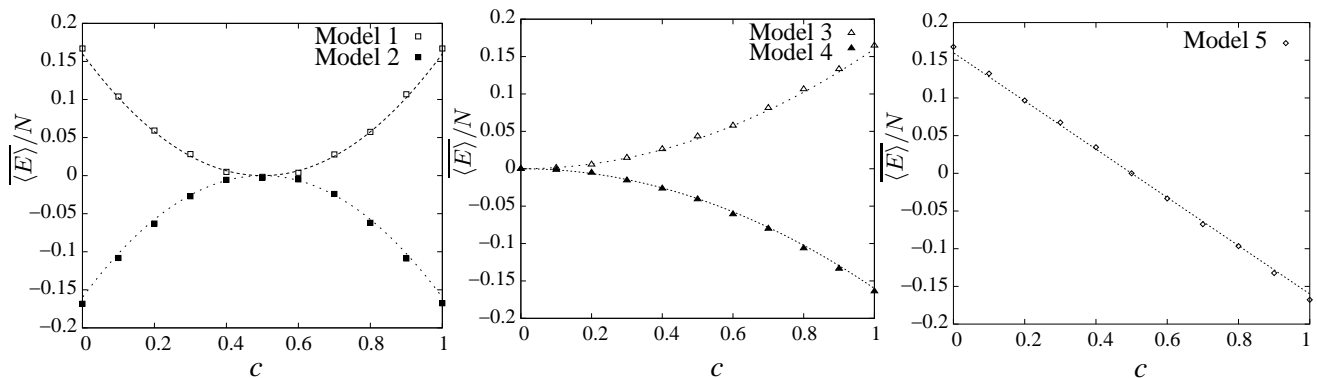


FIG. 1: The averaged energy per monomer of an  $N = 100$ -monomer heterogeneous chain in the high-temperature regime with various types of intramonomer interaction as function of  $c$ . Lines are the corresponding results of empirical estimations (15)-(19).

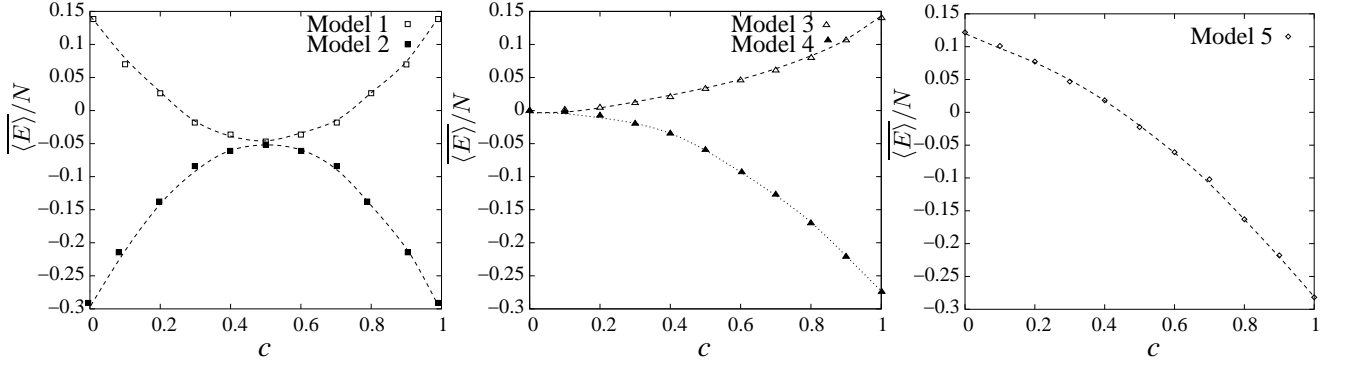


FIG. 2: The averaged energy per monomer of an  $N = 100$ -monomer heterogeneous chain with various types of intramonomer interaction as function of  $c$  at temperature  $T = 4.0$ . Lines are guides to the eye.

the number of  $BB$  nearest-neighbour contacts:

$$n_{BB} = \frac{1}{2}Np(1-c) = \frac{1}{2}Np(1-c)^2, \quad (13)$$

and the number of  $AB$  nearest-neighbour contacts:

$$n_{AB} = Nap(1-c) = Npc(1-c). \quad (14)$$

Thus, for the total energies of models 1-5 we have correspondingly:

$$E^{\text{Model 1}} = n_{AA} + n_{BB} - n_{AB} = 2Np \left( c^2 - c + \frac{1}{4} \right), \quad (15)$$

$$E^{\text{Model 2}} = -n_{AA} - n_{BB} + n_{AB} = -2Np \left( c^2 - c + \frac{1}{4} \right), \quad (16)$$

$$E^{\text{Model 3}} = n_{AA} = \frac{1}{2}Npc^2, \quad (17)$$

$$E^{\text{Model 4}} = -n_{AA} = -\frac{1}{2}Npc^2, \quad (18)$$

$$E^{\text{Model 5}} = -n_{AA} + n_{BB} = Np \left( \frac{1}{2} - c \right). \quad (19)$$

In the high-temperature regime, the steric self-avoidance effect between monomers plays the only role in determining the conformational properties of polymers. Thus taken the known value  $z=4.68404(9)$  [28] (which gives  $p = 0.31596(9)$ )

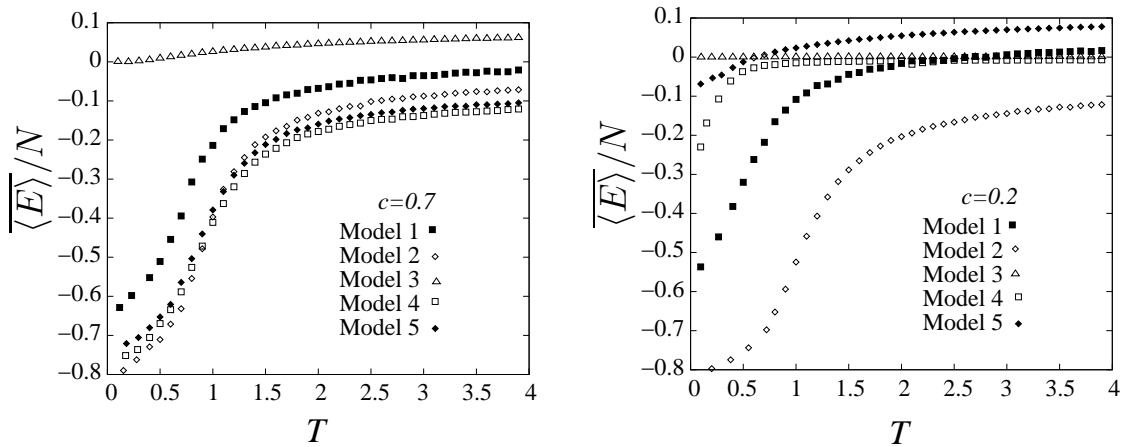


FIG. 3: Averaged energy per one monomer of the heterogeneous polymer chain with various types of monomer-monomer interactions as function of temperature at fixed inhomogeneity ratio  $c = 0.7$  and  $c = 0.2$ .

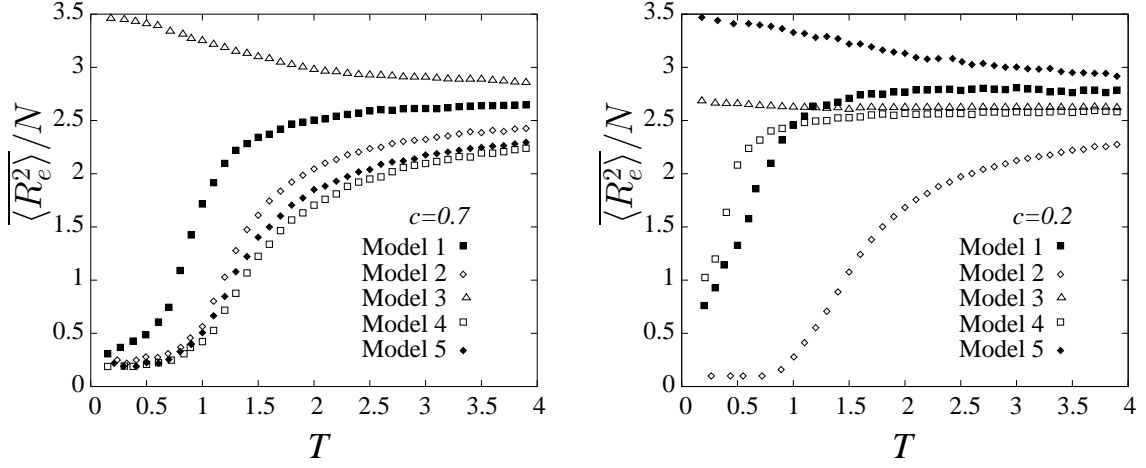


FIG. 4: Averaged end-to-end distance of the heterogeneous polymer chain with various types of monomer-monomer interactions as function of temperature at fixed inhomogeneity ratio  $c = 0.7$  and  $c = 0.2$ .

according to (11)), we can obtain numerical estimates for (15)-(19). These expressions are perfectly confirmed by our numerical simulations (see Fig. 1), where the corresponding energies per monomer are presented.

With lowering the temperature, the interactions between monomers become essential, and the  $p$ -values for each model appear to be dependent on the inhomogeneity ratio  $c$ . As an example, we present our simulation results for corresponding energies per monomer at  $T = 4.0$  in Fig. 2. Though for all models the polymer chains are expected to be still in an extended state (coil regime) at this temperature, the typical configurations they attain at given inhomogeneity ratio are governed by that nearest-neighbour contact value which minimizes the total energy. Indeed, for model 3 the most preferable configurations would be those with smaller value of nearest-neighbour contacts, whereas for model 4 the larger possible value of  $p$  will minimize the total energy. The values of  $p$  for all the models at various  $c$ -values can be estimated on the basis of data presented in Fig. 2 applying expressions (15)-(19).

In Fig. 3 we present our results for the averaged energies of the  $N = 100$ -monomer heterogeneous chain as function of temperature at fixed inhomogeneity ratios  $c = 0.7$  and  $c = 0.2$ . For  $c = 0.7$ , the energy of the models 1, 2, 4 and 5 decreases monotonically as temperature is lowered and sharply falls down within a short temperature interval, indicating the transition from extended state (coil regime) to compact state. For model 3, however, one observes only a monotonic decrease of energy, which gradually tends to zero with lowering the temperature, and thus no transition occurs. Moreover, within the frames of model 3, decreasing the energy means decreasing the averaged number of nearest-neighbour contacts, and thus expanding the chain. Thus, unlike the other models, in this case the

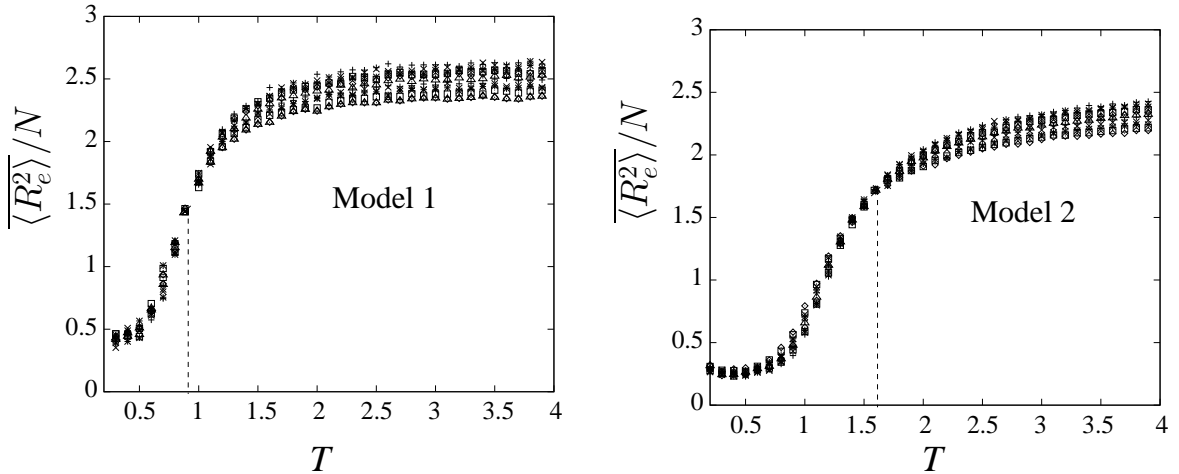


FIG. 5: Averaged end to end distance of a heterogeneous polymer chain divided by number of monomers as function of temperature at inhomogeneity ratio  $c = 0.7$  at various values of  $N$  from 68 up to 100.

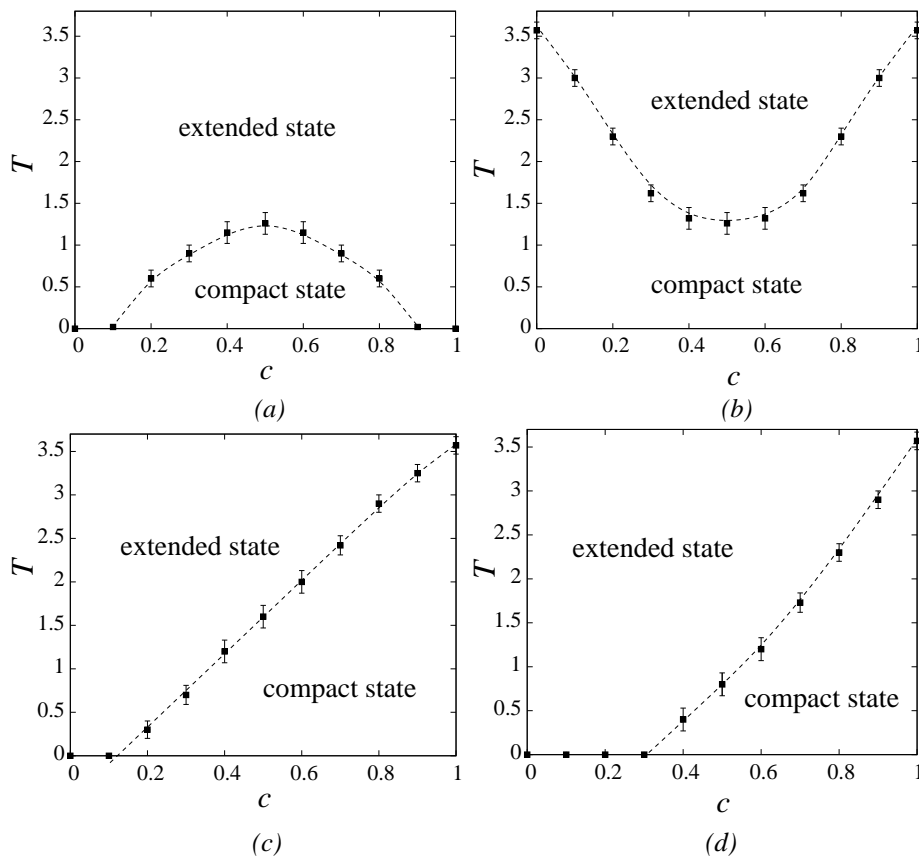


FIG. 6: Phase diagrams of heterogeneous polymer chain in  $T$ - $c$  space. (a) model 1, (b) model 2, (c) model 4, (d) model 5.

averaged size of the heteropolymer chain increases with lowering the temperature. Similar results are found at fixed inhomogeneity ratio  $c = 0.2$ . In this case, the model 5 consists mainly of monomers, which repel each other, and, like in model 3, conformations with a minimum number of nearest-neighbour contacts are energetically preferable, which leads to expanding the chain size with lowering the temperature. Note also that at small concentration of attractive monomers, the chains can attain the compact state only when they are long enough and have enough attractive nearest-neighbour contacts to overcome the conformational entropy.

In Fig. 4 our results for the averaged end-to-end distance of the  $N = 100$ -monomer heterogeneous chain as function of temperature are given at the same two fixed values of inhomogeneity ratio  $c = 0.7$  and  $c = 0.2$ . These results confirm our discussions above. For the case  $c = 0.7$ , one notices a decrease of the polymer chain size with lowering the temperature (models 1, 2, 4 and 5), which becomes crucial close to the transition temperature into the compact state. However, in model 3 the polymer chain is expanding its size with lowering the temperature due to repulsion between monomers, and the polymer chain remains in an extended state at any temperature. For  $c = 0.2$ , the size of polymer chain of model 5 increases even stronger than that of model 3 due to the considerably larger amount of mutually repelling monomers; in both cases the polymer chain remains in an extended state at any temperature.

To obtain a quantitative description of conformational transitions from extended to compact state in models 1-5, we study the ratio  $\langle R_e^2 \rangle / N$  for various  $N$  and fixed inhomogeneity ratios  $c$  as function of temperature (Fig. 5). Assuming that the collapse of heteropolymers is through the  $\Theta$ -transition similar to homogeneous polymers, we expect  $\langle R_e^2 \rangle \sim N$  at the  $\Theta$ -point, and thus the curves intersect in the vicinity of the transition temperature (the larger  $N$ , the steeper the curves). Despite the slight dependence of the intersection points on  $N$ , we can locate the crossover temperature. If at some  $T$  and  $c$  values the curves do not intersect, there is no phase transition and the polymer is in an extended state at any temperature.

As a result, we obtain the phase diagrams, presented in Fig. 6. Note, that within the frames of model 3 the polymer chain is always in an extended state at any temperature and inhomogeneity ratio value, thus we do not construct any phase diagram for this model. Models 2, 4 and 5 describe homogeneous polymer chains with nearest-neighbour attractions in the limiting case  $c = 1$  (for model 2 also  $c = 0$ ) with known value of transition temperature  $T_\Theta = 3.717(3)$  [26]. In model 2, the  $\Theta$ -transition is present always at any value of  $c$ , whereas models 1, 4 and 5

remain in the extended state when the concentration of attracting monomers is too small to cause a transition into the compact state. The results for the case of model 1 are in a good correspondence with those obtained previously in Ref. [25].

#### IV. CONCLUSIONS

We studied the conformational properties of heteropolymers within the frames of a lattice model containing  $N_A$  monomers of type  $A$  and  $N_B = N - N_A$  monomers of type  $B$  in  $d = 3$  dimensions. Restricting ourselves only to short-range interactions between any pair of monomers residing on neighboring lattice sites that are not connected by a covalent bond, we consider 5 different generalizations of this model (1)-(5). In particular, the model 1 where the like monomers repel and the opposite ones attract, refers to strongly screened Coulomb interactions, as studied previously in Ref. [25]. The model 3 is a particular case of model 1 and corresponds to a polymer chain containing charged ( $A$ ) and neutral ( $B$ ) monomers. Model 4 refers to the (minimal) HP model [5] with hydrophobic ( $A$ ) and hydrophilic ( $B$ ) monomers. Cases 2 and 5 can be considered as generalizations of the two above mentioned models.

Applying the pruned-enriched Rosenbluth chain-growth algorithm (PERM) we analyze numerically the peculiarities of transitions from extended into compact state as function of inhomogeneity ratio  $c \equiv N_A/N$  for all the heteropolymer chain models. Within the frames of model 3, unlike the other models, the polymer chain is expanding its size with lowering the temperature due to repulsion between monomers, and the polymer chain remains in an extended state at any temperature. In model 2, the  $\Theta$ -transition is present always at any value of inhomogeneity ratio  $c$ , whereas models 1, 4 and 5 remain in an extended state when the concentration of attracting monomers is too small to cause a transition into the compact state. Note also that at small concentration of attractive monomers, the chains can attain the compact state only when they are long enough and have enough attractive nearest-neighbour contacts to overcome the conformational entropy. Models 2, 4 and 5 describe homogeneous polymer chains with nearest-neighbour attractions in the limiting case  $c = 1$  (for model 2 also  $c = 0$ ) with known value of transition temperature  $T_\Theta = 3.717(3)$  [26].

#### Acknowledgements

This work was supported by an Institute Partnership grant of the Alexander von Humboldt Foundation. W.J. acknowledges support by DFG Sonderforschungsbereich SFB/TRR 102 (Project B04) and Sächsische DFG Forschergruppe FOR877 under Grant No. JA 483/29-1.

- 
- [1] P.G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1979).
  - [2] J. desCloizeaux and G. Jannink, *Polymers in Solution: Their Modeling and Structure* (Clarendon Press, Oxford 1990).
  - [3] A. Yu. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* (AIP Press, New York, 1994).
  - [4] N. Clisby, Phys. Rev. Lett. **104**, 055702 (2010).
  - [5] K.A. Dill, Biochemistry **24**, 1501 (1985); K.F. Lau and K.A. Dill, Macromolecules **22**, 3986 (1989).
  - [6] F.H. Stillinger, T. Head-Gordon, and C.L. Hirshfeld, Phys. Rev. E **48**, 1469 (1993).
  - [7] F.H. Stillinger and T. Head-Gordon, Phys. Rev. E **52**, 2872 (1995).
  - [8] A. Irbäck, C. Peterson, and F. Potthast, Phys. Rev. E **55**, 860 (1997).
  - [9] A. Irbäck, C. Peterson, and F. Potthast, and O. Sommelius, J. Chem. Phys. **107**, 273 (1997).
  - [10] H.-P. Hsu, V. Mehra, and P. Grassberger, Phys. Rev. E **68**, 037703 (2003).
  - [11] M. Bachmann and W. Janke, J. Chem. Phys. **120**, 6779 (2004).
  - [12] M. Bachmann, H. Arkin, and W. Janke, Phys. Rev. E **71**, 031906 (2005).
  - [13] S. Schnabel, M. Bachmann, and W. Janke, J. Chem. Phys. **126**, 105102 (2007).
  - [14] C. Tanford, *Physical Chemistry of Macromolecules* (Wiley, New York, 1961).
  - [15] For a recent review, see e.g.: A.V. Dobrynin, R.H. Colby, and M. Rubinstein, J. Polym. Sci. B: Polymer Physics **42**, 3513 (2004).
  - [16] Y. Kantor, H. Li, and M. Kardar, Phys. Rev. Lett. **69**, 61 (1992); Y. Kantor, M. Kardar, and H. Li, Phys. Rev. E **49**, 1383 (1994).
  - [17] A.M. Gutin and E.I. Shakhnovich, Phys. Rev. E **50**, R3322 (1994).
  - [18] Y. Kantor and M. Kardar, Phys. Rev. E **52**, 835 (1995).
  - [19] A.V. Dobrynin, M. Rubinstein, and S.P. Obukhov, Macromolecules **29**, 2974 (1996).
  - [20] M.C. Barbosa and Y. Levin, Physica A **231**, 467 (1996).
  - [21] M. Tanaka, A.Yu. Grosberg, V.S. Pande, and T. Tanaka, Phys. Rev. E **56**, 5798 (1997).

- [22] R. Everaers, A. Johner, and J.-F. Johny, *Europhys. Lett.* **37**, 275 (1997), *Macromolecules* **30**, 8478 (1997).
- [23] A.V. Lyubin, B. Dünweg, O.V. Borisov, and A.A. Darinskii, *Macromolecules* **32**, 3264 (1999).
- [24] V. Yamakov, A. Milchev, H.J. Limbach, B. Dünweg, and R. Everaers, *Phys. Rev. Lett.* **85**, 4305 (2000).
- [25] Y. Kantor and M. Kardar, *Europhys. Lett.* **28**, 169 (1994).
- [26] P. Grassberger, *Phys. Rev. E* **56**, 3682 (1997).
- [27] H.P. Hsu, V. Mehra, W. Nadler, and P. Grassberger, *J. Chem. Phys.* **118**, 444 (2007).
- [28] D. MacDonald, S. Joseph, D.L. Hunter, L.L. Moseley, N. Jan, and A.J. Guttmann, *J. Phys. A* **33**, 5973 (2000).